

shWiiFit Reduce Dependency Parsing

浅原 正幸[†]

本稿では係り受け構造情報のタグ付けの一貫性について考える。係り受け構造には、統語的制約により一意に決まる構造と選択選好性によるタグ付け作業者に委ねる構造がある。多くの場合、統語的制約を優先してタグ付けられるが、選択選好性に影響され誤ってタグ付ける例が多々ある。このような事例について誤り傾向の差分を評価するために、ゲームを用いた新しい心理言語実験手法を提案する。埋め込み構造によるガーデンパス文を用いて13人の被験者で実験を行ったほか、6種類の係り受け解析器を用いて解析誤り傾向の比較を行った。さらに最も誤った種類の文に対し、選択選好性がどのように影響したかについて報告する。

キーワード：係り受け解析，コーパス，統語的制約，選択選好性

shWiiFit Reduce Dependency Parsing

MASAYUKI ASAHARA[†]

This paper argues about consistency of syntactic dependency corpus annotation. Dependency structure involves unambiguous substructures specified by syntactic constraints and underspecified substructures evaluated by selectional preferences of annotators. In most cases, the syntactic constraints have a priority over the selectional preferences. In some cases, however, the selectional preferences are given priority over the syntactic constraints. We propose a new psycholinguistic experimental environment to use a fitness game application, in which a user stands on Nintendo Wii Balance Board and plays a parsing game competing on the accuracy and the speed. The game-based experiments investigate how correctly a human can parse written sentences. We evaluate syntactically confusing sentences with 13 experimental subjects and 6 dependency parsers. Moreover, we investigate effects of selectional preferences in the incorrectly annotated sentences.

Key Words: *Dependency Parsing, Corpus, Syntactic Constraint, Selectional Preference*

1 はじめに

言語解析器の作成時、タグ付きコーパスを用いて構造推定のための機械学習器を訓練する。しかし、そのコーパスはどのくらい一貫性をもってタグ付けられるものだろうか。一貫性のないコーパスを用いて評価を行うとその評価は信頼できないものとなる。また、一貫性のないコーパスから訓練すると、頑健な学習モデルを利用していたとしても解析器の性能は悪くなる。

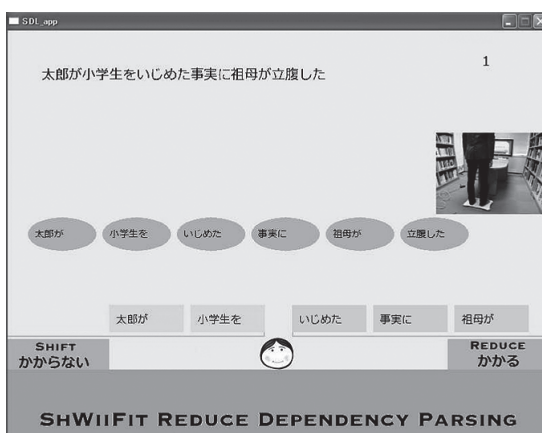
[†] 奈良先端科学技術大学院大学, Nara Institute of Science and Technology, Japan

本稿では、人間による日本語係り受け関係タグ付け作業に関して、どのくらい一貫性をもって正しくタグ付け可能かを評価する新しいゲームアプリケーション“shWiiFit Reduce Dependency Parsing”(図1)を提案する。ゲームのプレイヤーはWii バランスボードの上に立ち、係り受け解析対象の文を読み、画面中央の文節対に対して2種類の判断「係らない (SHIFT)」もしくは「係る (REDUCE)」の判断を選択し、体重を左右のどちらかに加重する。

係り受け構造のタグ付けにおける非一貫性は次の3つに由来すると考える。1つ目は、係り受け構造が一意に決まるが、作業者が誤るもの。2つ目は、複数の可能な正しい構造に対して、基準により一意に決めているが、作業者が基準を踏襲できなかったもの。3つめは、複数の可能な正しい構造に対して、基準などが決められていないもの。

ここでは、1つめの非一貫性つまりタグ付けの正確性について検討する。このゲームアプリケーションを用いて、埋め込み構造に基づくガーデンパス文(図2)のタグ付け困難性を評価する心理言語実験を行う。対象となる文は統語的制約のみによりその係り受け構造が一意に決定できる。しかしながら、被験者は動詞の選択嗜好性によるバイアスにより係り受け構造付与を誤ってしまう傾向があり、本稿ではその傾向を定量的に調査する。また、同じガーデンパス文を、各種係り受け解析器で解析し、現在の係り受け解析モデルの弱点について分析する。

人間の統語解析処理については、自己ペースリーディング法・質問法・視線検出法などの手法により心理言語学の分野で調査されてきた(Mazuka, Itoh, and Kondo 1997; Tokimoto 2004)。しかしながら、これらの心理言語学で用いられてきた手法は、読む速度を計測したり、文の意味を質問により事後確認したりする手法であり、コーパスに対する係り受けのタグ付けに直接寄与しない。一方、提案する手法では人間の係り受け判断をより直接的に評価し、また体重加



<http://goo.gl/XDTEb>

図1 shWiiFit Reduce Dependency Parsing

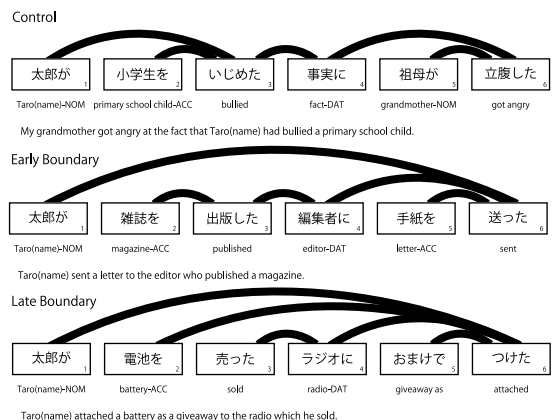


図2 Example Sentences

重分布に基づいて解析速度を追跡することができる。

以下、2節では日本語係り受け解析手法について概説する。3節では用いたガーデンパス文について説明する。4節では人間による係り受け解析の調査に用いたゲームについて紹介する。5節では実験結果と考察を示し、6節にまとめと今後の展開について示す。

2 日本語係り受け解析

日本語の係り受け解析器は京大コーパス由来のものが多く、コーパスの基準により、係り受け関係は文節単位に付与され、係り受け関係は交差することがほとんどなく (projective)、常に主辞が右にくる性質を持つ (strictly head-final)。また文節を単位とした場合、動詞は格フレームを持つがその格要素は頻繁に省略される (productive usage of empty category)、主辞に対する従属要素間の語順は比較的自由である (presence of scrambling) といった特性がある。このような特性により、他言語と比して係り受け解析手法を単純化することができる。

オープンソースの日本語係り受け解析器が2つある。1つは京都大学で開発された KNP であり、入力として規則に基づく形態素解析器 JUMAN の出力を用いる。KNP は同表記異義語の曖昧性解消・文節まとめあげ・並列構造解析・係り受け解析・格解析を行うことができ、文節まとめあげは規則に基づき、以降の処理は規則と生成モデルの混合手法に基づく。もう1つは工藤拓氏による CaboCha で、条件付確率場に基づく形態素解析器 MeCab の出力を用いる。CaboCha は文節まとめあげ・係り受け解析を行うことができ、それぞれ条件付確率場・サポートベクトルマシンといった識別モデルに基づく。

日本語の係り受け解析手法は決定的な状態遷移アルゴリズムに基づくものが多く提案されている。長さ N の文に対し、時間計算量 $O(N^2)$ の Cascaded Chunking 法 (Kudo and Matsumoto 2002)・ $O(N)$ の Shift Reduce 法 (Sassano 2004)・ $O(N^2)$ の Tournament 法 (Iwatate, Asahara, and Matsumoto 2008) などが提案されている。

他言語では様々な係り受け解析手法が提案されており (Buchholz and Marsi 2006; Nivre, Hall, McDonald, Nilsson, Riedel, and Yuret 2007)、グラフに基づく方法 (Eisner 2000; McDonald, Pereira, Ribarov, and Hajič 2005; Carreras 2007; Koo and Collins 2010) が決定的な状態遷移アルゴリズム (Nivre 2003; Nivre and Scholz 2004) とともに高精度を達成している。2手法が得意とする言語現象が異なるため、両者の組み合わせ手法 (Nivre and McDonald 2008) が提案されている。しかしながら、日本語の係り受け解析において、グラフに基づく手法が高精度を達成されたという報告はない。

3 日本語ガーデンパス文

ガーデンパス文とは、途中まで文字列を読んで一旦理解されやすい解釈が誤っており、最後まで読んで初めて正しい解釈ができるような構造を持つ文のことをいう。

本稿では係り受け構造同定を誤りやすい文として、Tokimoto (Tokimoto 2004) の実験で用いられた埋め込み構造に基づく日本語ガーデンパス文を用いる。利用する例文は以下の形式の6文節からなる：

$$NP_1^{NOM} \quad NP_2^{ACC} \quad V_3^{PAST} \quad NP_4^{DAT} \quad X_5 \quad V_6^{PAST}$$

下添え字は文節の順番を表す便宜上の数字である。NP は格助詞を持つ名詞句を表し、 NP^{NOM} はガ格を持つ名詞句・ NP^{ACC} はヲ格を持つ名詞句・ NP^{DAT} は二格を持つ名詞句を表す。 V^{PAST} はタ形の動詞を表す。

5 番目の文節 X に何が割り当てられるかによって異なる3種類の係り受け構造を持つ。1つ目はガ格を持つ名詞句 NP_5^{NOM} を割り当てたもので Control (CTRL) 文と呼ぶ。2つ目はヲ格を持つ名詞句 NP_5^{ACC} を割り当てたもので Early Boundary (EB) 文と呼ぶ。3つ目はそれ以外の句を割り当てたもので Late Boundary (LB) 文と呼ぶ。3種類の係り受け構造を図2に示す。

CTRL 文の場合、並列構造などの例外を除いて1つの動詞が2つ以上のガ格を持たないため、最初の NP_1^{NOM} が V_3^{PAST} のガ格、 NP_5^{NOM} が V_6^{PAST} のガ格となる。非交差条件と常に主辞が右に来る制約により NP_2^{ACC} が V_3^{PAST} に、 NP_4^{DAT} が V_6^{PAST} に係る。尚、 V_3^{PAST} が NP_4^{DAT} に連体節外の関係 (寺村 1981) で係る。

EB 文の場合、並列構造などの例外を除いて1つの動詞が2つ以上のヲ格を持たないため、最初の NP_2^{ACC} が V_3^{PAST} のヲ格、 NP_5^{ACC} が V_6^{PAST} のヲ格となる。準備した文は全て NP_4^{DAT} が V_3^{PAST} を連体節内の関係で受け、意味的には V_3^{PAST} のガ格に相当する。このため NP_1^{NOM} が V_6^{PAST} のガ格になる。

LB 文の場合、準備した文は全て NP_4^{DAT} が V_3^{PAST} を連体節内の関係で受け、意味的には V_3^{PAST} のヲ格に相当する。このため NP_2^{ACC} が V_6^{PAST} のヲ格になる。非交差条件と常に主辞が右に来る制約により NP_1^{NOM} が V_6^{PAST} に係る。

4 ゲーム：shWiiFit Reduce Dependency Parsing

本節では人間の係り受け解析判定を調査するために開発したゲーム “shWiiFit Reduce Dependency Parsing” について説明する。図1にゲーム画面を示す。

ゲームはNivre らの shift reduce 法 (Nivre 2003; Nivre and Scholz 2004) の日本語対応版である Sassano のアルゴリズムを元に行っている (Algorithm 1)。4種類のアクション “default reduce” ・ “default shift” ・ “REDUCE” ・ “SHIFT” のうち “REDUCE” ・ “SHIFT” を人間が判断する。文

Algorithm 1 Shift reduce 法に基づく日本語係り受け解析

```

 $\langle S, Q, A \rangle = \langle nil, W, \phi \rangle$  % Initialization
repeat
  if  $S! = nil$  and  $|Q| == 1$  then
     $\langle s|S, q, A \rangle \Rightarrow \langle S, q, A \cup \langle s, q \rangle \rangle$  % “default reduce”
  else if  $S == nil$  and  $|Q| > 1$  then
     $\langle nil, q|Q, A \rangle \Rightarrow \langle q, Q, A \rangle$  % “default shift”
  else
    if  $s$  and  $q$  has a dependency relation then
       $\langle s|S, q|Q, A \rangle \Rightarrow \langle S, q|Q, A \cup \langle s, q \rangle \rangle$  % “REDUCE”
    else
       $\langle s|S, q|Q, A \rangle \Rightarrow \langle q|s|S, Q, A \rangle$  % “SHIFT”
    end if
  end if
until  $S == nil$  and  $|Q| == 1$ 
return  $A$ 

```

節数 N の 1 文を，高々 $2N$ 回のアクションを決定することにより係り受け解析ができる．以下アルゴリズムについて詳しく説明する：

スタック S ・キュー Q ・解析済み係り受け関係を格納する A の 3 つ組を定義する．それぞれ，空スタック nil ・入力文節列 W ・空集合 ϕ で初期化する．係り受け解析はこの 3 つ組の状態遷移により進められる．状態遷移は以下の条件分岐により行う：

- “default reduce”：スタック S が空ではなくかつキュー Q が単一文節 q のみの場合，システムが自動的に S の先頭要素 s が q に係る関係を解析済み係り受け関係集合 A に追加し S から s を pop する．
- “default shift”： S が空でありかつ Q が複数文節からなる場合，システムが自動的に Q の先頭要素 q を pop し， S に q を push する．
- “REDUCE”：人間が S の先頭要素 s と Q の先頭要素 q の間の係り受け関係を判断し係り受け関係がある場合， S の先頭要素 s が q に係る関係を A に追加し S から s を pop する．
- “SHIFT”：人間が S の先頭要素 s と Q の先頭要素 q の間の係り受け関係を判断し係り受け関係がない場合， Q の先頭要素 q を pop し， S に q を push する．

以下，アクションをどのようにプレイヤーが入力するかを説明する．ゲームのプレイヤーには図 1 のようなスクリーンが提示される．解析すべき文がスクリーンの上部に示される．ゲーム開始時には顔アイコンがスクリーン下部中央に示される．その左上にあるトレイがスタック S に相当し，右上にあるトレイがキュー Q に相当する．左トレイ S には何も載っていない状態

で、右トレイ Q には文の先頭 3 文節が載っている状態で初期化される。ゲームの間左トレイ S ・右トレイ Q とともに画面中央を先頭として高々 3 文節がプレイヤーに提示される。プレイヤーは Nintendo バランス Wii ボード上に乗り、体重移動により顔アイコンを移動させ左右の壁に移動させることにより注目する 2 文節間の係り受け関係の有無を入力する。もし左トレイ S の先頭要素 s と右トレイ Q の先頭要素 q が係り受け関係にある場合、プレイヤーは右方向に体重移動することにより顔アイコンを画面下部右の“REDUCE”という壁に移動させる。係り受け関係がない場合、左方向に体重移動することにより顔アイコンを画面下部左の“SHIFT”という壁に移動させる。壁に顔アイコンが触れた時点で入力と見なされ、ゲームは 820–860 msec. のアニメーションとともに、対応するアクションを実行する。この間プレイヤーが体重を左右に加重していない状態であれば、自動的にアイコンは中央に戻る。アクションが一意に決まる場合、つまり“default reduce”・“default shift”相当の場合、同様に 820–860 msec. のアニメーションが提示する。1 文解析後プレイヤーには解析結果が正しかったかどうか提示される。プレイヤーはジャンプすることにより次の文の解析に進むことができる。

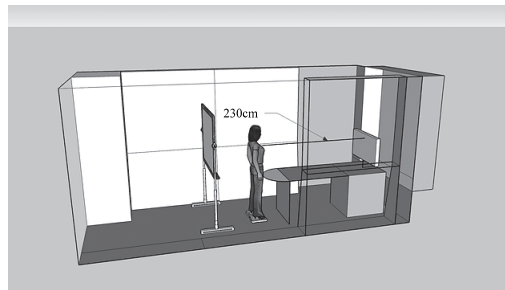
ゲームの間ソフトウェアは各アクションの反応時間を計測する。今回の実験では、入力デバイスとして、バランス Wii ボードを用いたが、ソフトウェアはキーボード上のカーソルキー・ジョイスティック・ゲームパッド (Nintendo Wii リモコンの各種センサを含む) でも動かすことができる。入力デバイスとしてバランス Wii ボードを用いた理由は、判断に迷った際にある程度プレイヤーに対して負荷を与えることができる点がある。プレイヤーが判断に迷った場合、キーボード上のカーソルキー・ゲームパッドを用いた場合にはプレイヤーは何もしなくてもよいが、バランス Wii ボードを用いた場合にはプレイヤーは体重を中心に保つ努力をしなければならない。この負荷の有無については、4 人に対する小規模の対照実験においてバランス Wii ボードの方が反応時間差が出やすいことを確認している。

5 実験

5.1 人間の係り受け解析判断

5.1.1 実験設定

4 節に示したゲームを用いて、心理言語実験を行う。ここでは 3 節に示した埋め込み構造に基づくガーデンパス文に対する係り受け解析の精度と反応時間について評価する。21–27 歳の日本語を母語とする大学生・大学院生 13 人を対象とし、被験者には謝金を支払う。ゲームの利用方法を教示するため、被験者は 6 ページのマニュアルを 5 分間読む。その後、係り受け関係の教示のために、文の構造を解説している小学校 4 年生の国語の教科書 2 ページを読む。教示時には速度と正解率の両方を計測していることを被験者に伝え、速く・正確に解析を行うことを依頼する。ゲーム中は各文が正解しているか否かのみを提示する。



<http://goo.gl/FP3G8>

図 3 実験環境

練習として被験者は3-6文節により構成される10文1セットを対象に実験を行う。被験者の希望により、同じ文を3セットまで練習することができる。練習に利用した平均文数は16.9文で5-12分要する。練習に用いた文の正しい係り受け構造は図入りのマニュアルに全て示されている。尚、13人の被験者のうちNintendo WiiFit および関連ゲームを高頻度で遊んだことのあるものは1人のみ。

本実験において60文を解析用データとして準備した。3節に示した判定対象となる3種類の例文(CTRL・EB・LB)10文ずつ計30文とフィラー30文からなる。判定対象となる文は全て6文節だが、フィラー文は6-7文節。判定対象となる文はTokimotoの実験で利用されたものと同じであり、フィラー文はTokimotoが行ったように文中に出現する語彙は10年分の新聞記事頻度とNTT語彙特性(天野, 近藤 1999)の語彙親密度により統制する。

各被験者は1回の本実験で40文を解析する。文は以下の順で提示する: 最初の5文はフィラー文・次の30文はフィラー15文+CTRL文5文+EB文5文+LB文5文をランダム順処理したもの・最後5文はフィラー文。被験者毎に異なるデータセットを作成する。1人の被験者に対して同じ文を2度提示しない。

図3に実験環境を示す。42インチのディスプレイ上には1024×768解像度で表示したうちの800×600ピクセルの部分がゲーム画面である。実験に利用した部屋は防音室ではないが、可能な限り静音化した。バランスボードと画面の距離は230cm。図中ホワイトボードの後ろに教示者がおり、練習時には操作方法の指示を行う。

5.1.2 実験結果

表1に文型毎の文正解率と文単位の反応時間の標準残差(内的にスチューデント化した残差)の平均と標準偏差を示す。標準残差の計算時は正しく解析した際の時間を以下のパラメータで

線形回帰する¹: (a) モーラ数・(b) 文字数・(c) 文節数・(d) 文の提示順・(e) “default reduce” の数・(f) “default shift” の数・(g) “REDUCE” の数・(h) “SHIFT” の数・(i) 隣接する 2 アクションで “SHIFT”/“REDUCE” が交替した回数 (逆順も含む)。

文正解率より, ガ格は文末の動詞に係け, ヲ格は近い動詞に係ける EB 文の構造をより正解できる傾向が見られる. 京大コーパスのマニュアルには, 格要素が両方に係る場合, ガ格は最右要素, ヲ格は最左要素に係けるとある. その点について被験者については教示していない. ゲームが shift reduce 法に基づくために全ての要素をより近い要素に係ける傾向があると予測したが, CTRL 文が EB 文に比べてより誤ることから必ずしも近い要素に係ける傾向があるとはいえない.

文単位の反応時間より, LB 文は他の文に比べて解析に時間がかかることがわかった (調整化残差の分析より 5%水準で有意差あり). CTRL 文と EB 文については統計的な有意差は見られなかった. 表 2 にアクション毎の反応時間を示す. 表の先頭行にアクションのインデックスを示す. 対象となる文は全て 6 文節からなり, 解析に必要なアクションの数は 10 であり, 表中の

表 1 文正解率と文単位の反応時間 (人間の係り受け解析判断)

		フィラー	CTRL	EB	LB
文正解率 (%)	平均	74	66	78	48
	標準偏差	19	19	39	24
反応時間標準残差	平均	-0.13	0.08	0.04	0.72
	標準偏差	0.11	0.67	0.57	0.45

表 2 アクション単位の反応時間 (人間の係り受け解析判断)

アクション順	0	1	2	3	4	5	6	7	8	9
CTRL	s	S	R	R	s	R	s	S	r	r
反応時間標準残差 (平均)		1.07	-0.06	-0.34		-0.13		-0.55		
反応時間標準残差 (標準偏差)		1.08	0.72	0.81		0.49		0.71		
実時間 (平均)	0.83	7.81	3.71	2.77	0.85	2.59	0.85	2.50	0.85	0.85
EB	s	S	R	S	R	S	S	r	r	r
反応時間標準残差 (平均)		0.98	0.15	-0.66	-0.16	-0.24	-0.07			
反応時間標準残差 (標準偏差)		1.06	0.94	0.74	0.73	0.73	0.84			
実時間 (平均)	0.83	7.71	3.83	2.39	2.60	3.36	3.64	0.85	0.85	0.85
LB	s	S	S	R	S	S	r	r	r	r
反応時間標準残差 (平均)		0.40	0.06	-0.19	-0.52	0.14				
反応時間標準残差 (標準偏差)		1.17	1.01	0.68	0.77	0.97				
実時間 (平均)	0.83	10.35	5.00	2.24	3.57	3.18	0.85	0.85	0.85	0.85

¹ 線形回帰は文の処理時間を正規化するために行う. 回帰による予測値と実測値の乖離を検討し, 所要時間の文型間の差異を評価する.

列はこれに対応する。このうちほぼ半分のアクションは “default reduce” ・ “default shift” である。表中には、各 3 文型 (CTRL・EB・LB) 毎の結果を 4 行で示し、1 行目は正しく解析するのに必要なアクション・2 行目は反応時間（アクション間の実時間差からアニメーション提示時間を引いたもの）標準残差の平均・3 行目は反応時間標準残差の標準偏差・4 行目はアクション間の実時間差の平均を示す。1 行目のアクションの略号はそれぞれ：“r” は “default reduce” ・ “s” は “default shift” ・ “R” は “REDUCE” ・ “S” は “SHIFT” を表す。2 行目と 3 行目の標準残差は正しく解析できた文についてアクション毎の反応時間を (a) モーラ数・(b) 文字数・(d) 文の提示順で線形回帰をしたものである。尚、文の反応時間のときに考慮した (c) および (e)-(i) は、各文型集合毎で全く一致するので考慮しない。

全ての文型について人間が反応可能な最初もしくは 2 番目のアクションで時間を要している。このことから、被験者は正しく解析するために、逐次的に解析するわけではなく、最初に全文を読んでから解析（の入力）をしていると考える。

5.1.3 関連研究

Mazuka らは様々な心理言語実験手法について紹介している (Mazuka et al. 1997)。ここでは各実験手法との対比を行う。

視線検出法では、被験者が文を読む際にどのくらい早く読むかを計測する。1 回目の走査の時間・読み直しを行うかどうか・2 回目の走査にかかる時間を計測する。この方法は文を読む速度を自然な方法で計測可能だが、視線検出システムは大変高価である。

自己ペースリーディング法は被験者の意思に基づいて文節（もしくは句）を逐次的に提示しながら、文を読む速度を計測する。文節を逐次的に提示するために、被験者は近い文節に係けるバイアスを持ち、ガーデンパス文を読む際には再解析 (reanalysis steps) を強いることになる。実験では、再解析コストを誤り率と文を読む速度により評価を行うことを目的とする。

質問法は被験者に対して文を理解しているか否かを直接聞く方法である。目的に応じて質問は統制される。*Who-did-what* 質問文では被験者が各文の意味を理解しているかどうかを評価する。*Difficulty rating* 質問文では被験者が各文を理解するのにどのくらい簡単もしくは困難だったかを評価する。*Misleadingness rating* 質問文では被験者が各文に対してどのくらい誤解しやすいかを評価する。

質問法は読む速度が得られる手法との組み合わせで行われることが多い。Tokimoto は図 2 に示す文を自己ペースリーディング法による読む速度の残差および *Who-did-what* 質問法による誤り率で評価した。彼の実験では、各文型の再解析コストは CTRL < EB < LB の順であった。我々の結果を誤り率に変換すると EB < (フィラー <) CTRL < LB となり結果が異なる。ガーデンパス効果は LB 文にのみ認められた。この違いは手法と目的の違いに由来する。Tokimoto は文節を逐次的に提示し、人間が文の意味を正しく理解しているか否かを調査することを目的

とする．対照的に我々の実験では被験者は実験中常に文全体を見ることができ，係り受け解析が正しく行えるかを調査することを目的とする．

最後に，我々の実験設定において係り受け解析を shift reduce 法に基づいて解析することが適切かどうか考察する．ゲームで提示されるトレイ（スタックおよびキュー）は文脈情報として，それぞれ3文節ずつの窓幅の情報を与える．1文節目と3文節目 ($\langle NP_1^{NOM}, V_3^{PAST} \rangle$) および2文節目と3文節目 ($\langle NP_2^{ACC}, V_3^{PAST} \rangle$) の係り受け関係を判断する際に6文節目の動詞 V_6^{PAST} の情報はトレイには表示されない．しかしながら画面の上部には常に文の全体を表示しており，この shift reduce 法に由来するバイアスを低減している．尚，文全体を表示しない（つまり文末の文節を見せない）設定で実験を行った場合，全て CTRL 文の構造に割り当て Tokimoto らの方法と同じく再解析に陥ることが，被験者5人による事前実験によりわかっている．

5.2 人間と係り受け解析器の比較

5.2.1 実験設定

5.1 節の心理言語実験で用いた 60 文を各種係り受け解析器で解析することにより，人間による結果と6種類の係り受け解析器の結果とを比較する．KNP-3.01 は形態素解析器 JUMAN-6.0 の出力を元とし，CaboCha-0.60pre4 は形態素解析器 MeCab-0.98+IPADIC-2.7.0 の出力を元とする．この2つの解析器についてはデフォルトのパラメータ設定を用いる．さらに，Shift Reduce 法 (Sassano 2004) に基づく実装を京大コーパス約 8,000 文で訓練したもの (Shift Reduce 8K) と約 34,000 文で訓練したもの (Shift Reduce 34K) の2つと，Tournament 法 (Iwatate et al. 2008) に基づく実装を約 8,000 文で訓練したもの (Tournament 8K) と約 34,000 文で訓練したもの (Tournament 34K) の2つを用いる．この4つの解析器では形態素解析器の出力として JUMAN-6.0 を，さらに正しい文節区切りを与え，機械学習器としてサポートベクトルマシンを用いる．学習器に与える素性については元の論文に可能な限り合わせた．

5.2.2 実験結果

表3に人間の判断と6種類の解析器の判断に基づく文正解率を示す．表中 [J] は Juman-6.0 の出力を，[M] は MeCab-0.98+IPADIC-2.7.0 の出力を用いていることを意味し，[GB] は正しい文節区切りを与えていることを意味する．フィラー列は一般的な文に対する性能を表す．

全ての解析器は CTRL 文において性能が良く LB 文においては性能が悪かった．この結果より，解析器は CTRL 文中の文節 $NP_1^{NOM} \cdot NP_2^{ACC} \cdot NP_4^{DAT}$ に対して正しい最も近い右要素 V_*^{PAST} に係けることができていたといえる．KNP-3.01 の CTRL 文に対する誤りは文節区切りの誤りであった．

CaboCha-0.60pre4 を含む shift reduce 法は近い係り受け関係を選好する傾向がある．実際， $\langle NP_1^{NOM}, V_3^{PAST} \rangle$ もしくは $\langle NP_2^{ACC}, V_3^{PAST} \rangle$ の係り受け関係を判断する際に，6番目の文

表 3 文正解率 (%) (人間と解析器の係り受け解析判断)

	フィラー	CTRL	EB	LB
人間の判断	74	66	78	48
KNP-3.01 [J]	83	90	60	50
CaboCha-0.60pre4 [M]	73	100	0	0
Shift Reduce 8K [J][GB]	70	100	0	0
Shift Reduce 34K [J][GB]	73	100	0	0
Tournament 8K [J][GB]	70	100	60	0
Tournament 34K [J][GB]	80	100	10	0

節 V_6^{PAST} は機械学習器の素性の窓幅の外側にあり評価されない。ゆえにこれらの解析器は EB 文と LB 文を全く解析できない。

Tournament 法は長い距離に係る係り受け関係をステップラダートーナメントにより考慮することができるため、いくつかの EB 文を解析できている。訓練データ量 8,000 文と 34,000 文との結果の差異は、係り受け関係 $\langle NP_1^{NOM}, V_3^{PAST} \rangle$ が 8,000 文データには出現しないが 34,000 文データには出現することによると考える。LB 文については、京都大学テキストコーパスのタグ付け基準「格要素が両方に係る場合、主格以外は近い方に係ける」が影響していると考ええる。

KNP-3.01 が EB・LB 文に関して最良の結果を達成している。KNP-3.01 は格解析モデル (Kawahara and Kurohashi 2006) を含んでおり、EB・LB 文の統語的意味的曖昧性のある程度解消できたのではないかと考える。

EB・LB 文を正しく解析するためには、共起の情報 (Abekawa and Okumura 2006)・格フレーム情報 (Kawahara and Kurohashi 2006)・格フレーム間遷移情報 (Kawahara and Kurohashi 2009) を用いる手法が考えられる。特に LB 文には次節で示すように生成モデルによる選択選好性のみでは解決できない文が含まれており、正しい解析のためには選択選好性と格要素の重複性排除の両方を同時推論する方策が必要であろう。

5.2.3 関連研究

Abekawa らは今回対象とした現象に似た関係節内の関係・外の関係の識別を行った (Abekawa and Okumura 2005)。本研究で扱った CTRL 文は外の関係に対応し、EB・LB 文は内の関係に対応する。彼らは決定木モデルを用いて、共起情報・格フレーム情報・外の関係になる度合いを用いて、この 2 種類を識別する手法を提案した。

5.3 Late Boundary 各文の分析

ここでは人間にも解析器にも解析が困難であった Late Boundary (LB) 文について、統語的制約と選択選好性の観点から分析する。LB 文は主節の動詞 V_6 がガ格ヲ格ニ格の三項を取り、関

係節内の動詞 V_3 がガ格ヲ格の二項を取る．関係節内の動詞 V_3 が後置する係り先の NP_4 をヲ格として取るために二重ヲ格違反と常に左に係る統語的制約により一意に構造が決まる．これに対し「より近い要素に係ける」という選好性と、主節の動詞 V_6 ・関係節内の動詞 V_3 間の選択選好性の強弱関係により、制約を無視してしまい誤るという傾向がある．より詳細に分析するために LB 文 10 文について、人の文正解率・KNP の結果・京都大学格フレームの頻度の大小関係・Google のヒット件数に基づく正規化頻度の大小関係を表 4 に示す．京都大学格フレームの頻度について、 NP_2 vs NP_4 は $NP_2^{ACC}V_3$ と $NP_4^{ACC}V_3$ で、 V_3 vs V_6 は $NP_2^{ACC}V_3$ と $NP_2^{ACC}V_6$ で、それぞれどちらの頻度が高いかを表す．尚、0 は両方の組み合わせが京都大学格フレームに登録されていなかったことを表す．Google の正規化頻度は、前者を NP_2^{ACC} と NP_4^{ACC} の件数で、後者を V_3 と V_6 の件数で正規化を行う．例文は人の文正解率の良い順に並べられ、下線は選択選好性が例文を正しく解析するために良い影響を与えるものを表す．

KNP が正解している文は人間にとっても解析が容易で、LB 文の人正解率が高い傾向にある．京都大学格フレームが例文の係り受け全てを被覆していないが、人と誤り傾向の相関があることがわかる．Google の正規化頻度は全例文を被覆しているが、係り受け関係ではなく連続文字列によるものなので、人間の誤り傾向と Google のヒット件数による選択選好性の間に矛盾があるところはいくつかある．実際、3つの項をとる述語に対してガ格・ヲ格・ニ格の順に述語から遠く置かれる傾向にあることを考えると直接比較できるものではない．しかしながら、 NP_2 と NP_4 のどちらが V_3 に係るかを評価した際に京都大学格フレームだけでなく、Google ヒット件数でも NP_2 の方が強い場合には、人も間違える傾向にあることがわかる．このことから選択選好性の強弱関係が統語的に制約に反している場合には人は係り受け構造同定を誤りやすいことがわかる．

表 4 LB 文の実験結果と選択選好性

例文	正解率		京大格フレーム頻度		Google 件数	
	人	KNP	NP_2 vs NP_4	V_3 vs V_6	NP_2 vs NP_4	V_3 vs V_6
花子がコーヒーをあたためたカップにたっぷりとそそいだ	75	<u>OK</u>	NP_4 - V_3	0	NP_2 - V_3	NP_2 - V_6
太郎がテレビを修理した自転車に静かにのせた	75	<u>OK</u>	NP_4 - V_3	NP_2 - V_3	NP_2 - V_3	NP_2 - V_6
優子が刺身を冷やしたお皿にいていねいに盛りつけた	71	<u>OK</u>	0	0	NP_4 - V_3	NP_2 - V_6
優子が食器を買ったテーブルにきれいにかがった	57	<u>OK</u>	0	0	NP_4 - V_3	NP_2 - V_3
陽子がソーセージをゆでたスパゲティーにすばやく加えた	45	<u>OK</u>	NP_4 - V_3	0	NP_2 - V_3	NP_2 - V_3
陽子がセロテープを破ったポスターにびったりと貼った	42	NG	0	NP_2 - V_6	NP_4 - V_3	NP_2 - V_6
次郎が大金を盗んだタンスにこっそり隠した	40	NG	NP_2 - V_3	NP_2 - V_3	NP_2 - V_3	NP_2 - V_6
太郎が電池を売ったラジオにおまけでつけた	33	NG	0	0	NP_4 - V_3	NP_2 - V_6
次郎がロープを切った丸太にしっかり巻きつけた	33	NG	NP_2 - V_3	NP_2 - V_3	NP_2 - V_3	NP_2 - V_6
花子が切手を集めたマッチ箱に大切にしまった	20	NG	0	0	NP_2 - V_3	NP_2 - V_3

解釈 1

解釈 2

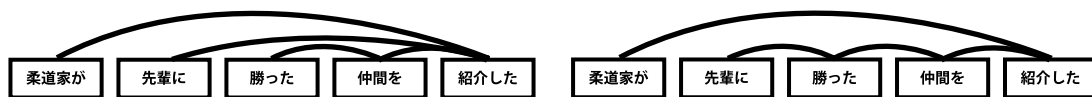


図 4 本質的に係り受け構造が曖昧な文

6 おわりに

本稿では係り受けタグ付けにおける人間の誤り傾向を調査する新しい方法を提案した。作成したゲームでは Nintendo バランス Wii ボードを入力デバイスとして、Shift Reduce 法に基づき、被験者がどのように係り受け解析を行うかを記録することができる。実験対象として埋め込み文に基づくガーデンバス文を用いて格要素の係り先の誤り傾向を観察した。また、同じ文を、6 種類の係り受け解析器で解析を行い、誤り傾向を比較した。さらに京都大学格フレームおよび Google のヒット件数に基づく選択選好性と誤り傾向の相関について調査した。調査結果より統語的制約のみで一意に文の係り受け構造が決まる場合においても、格構造の選択選好性の強弱関係により誤った係り受け構造を認識することがわかった。またその誤り傾向は従来の心理言語実験で行われている文の内容認識結果とは異なり、ヲ格を文末の述語に係ける文だけでなく、ガ格を関係節内の述語に係ける文でも係り受け構造同定をよく誤ることがわかった。本研究の今後の展開として真に統語的制約によって決まらない曖昧な構造の選択選好性の定量化がある。例えば図 4 のような本質的に係り受け構造が曖昧な文を考える：関係節内の動詞「勝った」はガ格とニ格を、主節の動詞「紹介した」はガ格とヲ格とニ格を取りうる項構造を持つ。「紹介した」が 3 つの項全てを埋める選好性（図 4 左）と「勝った」のニ格が省略されない選好性（図 4 右）とで、どちらを優先するかにより本質的に曖昧である。統語的制約によって決まらない構造を、選好性に基づき決定する過程は人によって揺れる。この本質的に曖昧な構造に対して、複数人の判定がどのように揺れるかを定量的に評価することにより、係り受けアノテーション情報の重層化を目指したい。アノテーションスキーマとして一致率を上げる展開とは別に、複数人のタグ付けが一致しない場合のその情報の保存という展開が考えられるのではないだろうか。

謝 辞

本研究の遂行にあたり千葉大学伝康晴氏と奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座の諸氏より助言をいただきました。ここに謝意を表します。

参考文献

- Abekawa, T. and Okumura, M. (2005). “Corpus-based Analysis of Japanese Relative Clause Constructions.” In *Proc. of the 2nd International Joint Conference on Natural Language Processing (IJCNLP-05)*, pp. 46–57.
- Abekawa, T. and Okumura, M. (2006). “Japanese Dependency Parsing Using Co-Occurrence Information a Combination of Case Elements.” In *Proc. of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-2006)*, pp. 833–840.
- Buchholz, S. and Marsi, E. (2006). “CoNLL-X Shared Task on Multilingual Dependency Parsing.” In *Proc. of the 10th Conference on Computational Natural Language Learning (CoNLL-X)*, pp. 149–165.
- Carreras, X. (2007). “Experiments with a Higher-order Projective Dependency Parser.” In *Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL-2007)*, pp. 957–961.
- Eisner, J. M. (2000). *Advances in Probabilistic and Other Parsing Technologies*, Chap. 1. Bilexical grammars and their cubic-time parsing algorithms. Kluwer Academic Publishers.
- Iwatate, M., Asahara, M., and Matsumoto, Y. (2008). “Japanese Dependency Parsing Using a Tournament Model.” In *Proc. of the 22nd International Conference on Computational Linguistics (COLING-2008)*, pp. 361–368.
- Kawahara, D. and Kurohashi, S. (2006). “A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis.” In *Proc. of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL-2006)*, pp. 176–183.
- Kawahara, D. and Kurohashi, S. (2009). “Capturing Consistency between Intra-clause and Inter-clause Relations in Knowledge-rich Dependency and Case Structure Analysis.” In *Proc. of the 11th International Conference on Parsing Technology (IWPT-2009)*, pp. 108–116.
- Koo, T. and Collins, M. (2010). “Efficient Third-Order Dependency Parsers.” In *Proc. of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1–11, Uppsala, Sweden. Association for Computational Linguistics.
- Kudo, T. and Matsumoto, Y. (2002). “Japanese Dependency Analysis using Cascaded Chunking.” In *Proc. of the 6th Conference on Natural Language Learning (CoNLL-2002)*, pp. 1–7.
- Mazuka, R., Itoh, K., and Kondo, T. (1997). “Processing down the garden path in Japanese: processing of sentences with lexical homonyms.” *Journal of Psycholinguistic Research*, **26**

(2), pp. 207–228.

McDonald, R., Pereira, F., Ribarov, K., and Hajič, J. (2005). “Non-Projective Dependency Parsing using Spanning Tree Algorithms.” In *Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT-EMNLP-2005)*, pp. 523–530.

Nivre, J. (2003). “An Efficient Algorithm for Projective Dependency Parsing.” In *Proc. of the 8th International Workshop on Parsing Technologies (IWPT 03)*, pp. 149–160.

Nivre, J., Hall, J., McDonald, S. K. R., Nilsson, J., Riedel, S., and Yuret, D. (2007). “The CoNLL 2007 Shared Task on Dependency Parsing.” In *Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL-2007)*, pp. 915–932.

Nivre, J. and McDonald, R. (2008). “Integrating Graph-Based and Transition-Based Dependency Parsers.” In *Proc. of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT-ACL-2008)*, pp. 950–958.

Nivre, J. and Scholz, M. (2004). “Deterministic Dependency Parsing of English Text.” In *Proc. of the 20th International Conference on Computational Linguistics (COLING-2004)*, pp. 64–70.

Sassano, M. (2004). “Linear-Time Dependency Analysis for Japanese.” In *Proc. of the 20th International Conference on Computational Linguistics (COLING-2004)*, pp. 8–14.

Tokimoto, S. (2004). “Reanalysis Costs in Processing Japanese Sentences with Complex NP Structures and Homonyms: Individual Differences and Verbal Working Memory Constraints.” Tech. rep. JCSS-TR-53, Japanese Cognitive Science Society.

天野成昭, 近藤公久 (1999). 日本語の語彙特性 第1期. 三省堂.

寺村秀夫 (1981). 日本語の文法 (下). 国立国語研究所.

略歴

浅原 正幸：2003 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。2004 年同大学助手，現在に至る。博士（工学）。

(2011 年 1 月 17 日 受付)

(2011 年 3 月 10 日 再受付)

(2011 年 4 月 12 日 採録)